ChinaXiv合作期刊

chinaXiv:202104.00064v1

泛基因组研究在遗传多样性和功能基因组学中的应用

向坤莉[1]，贺文闯[1]，邹益[1]，彭丹[1]，张晓妮[1]，廖雪竹[1]，王杰[1,2]，杨健康[3]，武志强[1*]

(1. 中国农业科学院农业基因组研究所，广东 深圳 518120; 2. 浙江农林大学 风景园林与建筑学院，杭州 311300; 3. 大理大学 基础医学院，云南 大理 671000)

摘要：相对于单个参考基因组仅聚焦于个体遗传信息的挖掘，泛基因组研究则能够反映整个物种或类群全部的遗传信息。随着基因组测序和分析技术的不断发展，泛基因组学逐渐成为新的研究热点，并已在植物、动物和微生物多个物种中获得了广泛应用，为全面解析物种或类群水平的遗传变异和多样性、功能基因组和系统进化重建等研究提供了强有力的工具，取得了很多显著的研究成果。尽管如此，由于泛基因组学研究尚处于发展阶段，测序费用和分析成本仍然较高，难以广泛普及；且存在分析标准不一、数据挖掘不够全面深入、理论难以应用于生产实际等尚待解决的问题，仍有较大的发展空间。该文系统总结了泛基因组在生物遗传多样性挖掘和功能基因组学中的研究进展，主要包括其在泛基因组图谱的构建、基因组变异和有利基因的发掘、功能基因的多态性、群体遗传多样性和系统进化等多个领域中的应用和研究，探讨了其在不同领域的应用潜力。同时，讨论了目前泛基因组研究中存在的局限性和可能的解决方法，并对其将来的发展前景进行了展望。

关键字：泛基因组，结构变异，功能基因，遗传多样性，系统进化

中图分类号：Q943.2　　　　　　　　　　**文献标识码**：A

Application of pan-genome in genetic diversity and functional genomics

XIANG Kunli[1], HE Wenchuang[1], ZOU Yi[1], PENG Dan[1], ZHANG Xiaoni[1], LIAO Xuezhu[1], WANG Jie[1,2], YANG Jiankang[3], WU Zhiqiang[1*]

(1. *Shenzhen Agricultural Genome Research Institute*, *Chinese Academy of Agricultural Sciences*, Shenzhen 518120, Guangdong, China; 2. *School of Landscape and Architecture*, *Zhejiang Agriculture & Forestry University*, Hangzhou 311300, China; 3. *School of Basic Medical Sciences*, *Dali University*, Dali 671000, Yunnan, China)

Abstract: The pan-genome can represent all of the genetic diversities in a species or population, which is a limitation for obtaining only one single reference genome. The pan-genomics is becoming a new hot research area and being widely applicated in researches of many species in plants, animals and microorganisms, as the development of the whole genome sequencing and analysis technology. It provided powerful tools for resolving the genetic variation and polymorphism at levels of species or taxa, researches of functional genomics and reconstruction of phylogenetics, obtaining abundant of significant research achievements. However, present researches on pan-genomics still need improve due to several problems, e.g., extensive cost of whole genome sequencing and data analysis, inconsistent analysis standards, lack of deeper and comprehensive explanation of the obtained data, and difficulty of application of the research

---

基金项目：国家自然科学基金（31970244）；中国农业科学院深圳农业基因组研究所启动资金［Supported by the National Natural Science Foundation of China (31970244); Initial fund of Shenzhen Agricultural Genome Research Institute, Chinese Academy of Agricultural Sciences］。

作者简介：向坤莉(1991-)，博士后，主要从事植物进化方面的研究，（E-mail）kunlixiang@caas.cn。

*__通信作者__：武志强，博士，研究员，主要从事植物进化基因组研究，（E-mail）wuzhiqiang@caas.cn。

achievements. We summarized the research progresses of pan-genomes on exploitation of genetic diversity and functional genomics, including construction of a pan genome map, identification of genome variations and favorable genes, polymorphism of functional genes, population genetic diversity and systematic evolution, and discussed its potential in application of different research fields. Furthermore, we discussed the limitations existed in the present studies and possible solutions, and directed the development prospect in the future on pan-genomics.

Key words: Pan-genome, structural variants, functional gene, genetic diversity, systematic evolution

chinaXiv:202104.00064v1

遗传变异是生物进化的内在源泉，因而，研究遗传多样性及其演化规律是生物遗传学及进化生物学研究中的核心问题之一。而泛基因组研究则是近年来随着测序成本的急剧降低和分析技术的快速发展而全面反映物种遗传变异的一种新兴工具。泛基因组研究能够从物种或类群水平广泛发掘和利用遗传变异多样性，是现代医学、生物学、农学中的一个前沿领域。其中，泛基因组(pan-genome)指一个物种或者类群的全部基因组信息的集合，包括核心基因组(core genome)和非必须基因组(dispensable genome)两部分。核心基因指在所有个体中都存在的基因/组分集合；而非必须基因组是指，在部分个体或单个个体中存在的基因/组分集合，有时也称为可变基因组(variable genome) (图 1; Tettelin et al., 2005; Medini et al., 2005)。核心基因组由所有样本中都存在的序列组成，往往与重要的生物学功能和表型特征相关，多数是一些管家基因(house-keeping genes)，反映了物种的稳定性；可变基因组由仅在部分样本中存在的序列组成，一般与物种对特定环境的适应性或特有的生物学特征相关，反映了物种的多样性和特异性(Montenegro et al., 2017; Gordon et al., 2017; Wang et al., 2018; Zhao et al., 2018; Liu et al., 2020)。

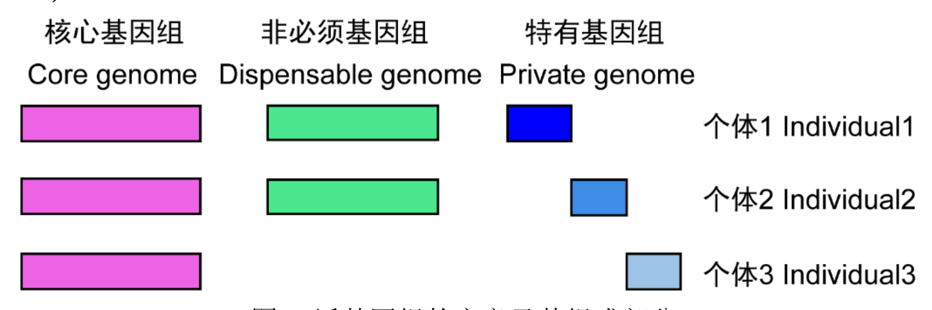


图 1 泛基因组的定义及其组成部分

Fig. 1 Definition and components of the pan-genome

当前，泛基因组研究已经广泛应用于多个植物、动物和微生物物种中，为全面解析物种或类群水平的遗传变异、功能基因研究和系统进化重建等研究提供了强有力的工具，取得了很多显著的研究成果(付静和秦启伟, 2012; 王娅丽等, 2019; Tian et al., 2019; Chen et al., 2020; Dom ńguez et al., 2020; Weissensteiner et al., 2020; Liu et al., 2020)。然而，现有的泛基因组学研究主要聚焦于不同个体基因组序列和基因结构的变异(Montenegro et al., 2017; Zhao et al., 2018; Gao et al., 2019; Liu et al., 2020)，但对这些变异是如何介导基因结构和功能的改变，最终影响生物表型，以及这种遗传差异是如何与环境因子互作的，未能进行深入探讨。本文综述了泛基因组学在不同物种中的研究进展，对其在群体基因组变异、功能基因的鉴定和发掘、群体遗传多样性和系统进化等多个领域中的应用与研究进行了系统性总结，并对其应用前景和局限性进行了探讨。

1 泛基因组图谱的构建

最早在 2005 年，Tettelin et al.(2005)在对几种链球菌属细菌(GBS, group B *Streptococcus*)

的遗传多样性研究中提出微生物的泛基因组概念，指出核心基因组是在所有菌株中都存在的基因；非必须基因组（可变基因组）是仅在部分菌株中存在的基因。其中 GBS 菌共有的核心基因组占 80%，剩余 20%的基因组信息为非必须基因组。随后，2010 年 Li et al. (2005)通过对多个人类个体基因组进行组装和比较基因组学分析，提出了“人类泛基因组”的概念，也就是人类群体基因组信息的总和，并从中鉴定获得新发现的序列达到 19~40 Mb。而随着千人基因组计划的提出和实施，泛基因组在人类疾病方面的研究取得了许多重大突破，为精准医疗计划提供了可能(1 000 Genomes Project Consortium, 2012)。

之后，随着越来越多的物种完成了高质量基因组参考序列的组装，多个动植物物种中相继报道了泛基因组图谱的构建相关研究工作。例如，通过对全球 12 个种猪品种的基因组进行高质量组装，构建了猪的泛基因组图谱，发现中国的猪品种有大约 9 Mb 的泛序列（pan-sequences）与欧洲的猪品种存在差异，其中包括脂肪细胞脂解的必要调节因子 *TIG3* (Tazarotene-induced gene 3) (Tian et al., 2019)；对 19 个小麦品种的泛基因组分析发现，平均每个样本中含有 128 656 个基因，核心基因有 89 795 个(Montenegro et al., 2017)；利用 725 个番茄品种的基因组信息构建的番茄泛基因组图谱中，整个番茄泛基因组共包含 40 396 个基因，其中 74.2%是核心基因(Gao et al., 2019)。此外，泛基因组在水稻(Schatz et al., 2014;Yao et al., 2015; Sun et al., 2017; Wang et al., 2018; Zhao et al., 2018; Zhou et al., 2020)、大豆(Li et al., 2014; Liu et al., 2020; 祝光涛和黄三文, 2020)、玉米(Hufford et al., 2012; Hirsch et al., 2014; 简银巧等, 2017)等重要的植物物种均获得了广泛应用(表 1)。因此，构建整个物种的泛基因组图谱已成为广泛应用的基因组学方法，不仅能够发现全面的遗传信息，而且能为从物种和群体水平进行功能基因组学、系统进化和其他生物学研究提供更强有力的工具。

2 泛基因组学研究中序列结构变异与功能基因发掘

同一物种内一个或几个参考基因组能够反映的遗传变异是非常有限的，而泛基因组研究能够覆盖物种或类群中的所有变异，为研究整个物种或类群水平上的基因组序列和结构变异提供了可能。现代生物基因库中的遗传变异通常包括单核苷酸多态性(SNPs, single-nucleotide polymorphisms)、插入缺失(Indels, Insertions and deletions)和大的结构变异(SVs, large structural variants)。其中 SVs 主要包括拷贝数变异(CNVs, copy number variants)、存在/缺失变异(PAVs, presence/absence variants)、移位(translocation events)和倒置(inversion events)等，而这些变异往往和一些关键的农艺性状相关(Springer et al., 2009; Hirsch et al., 2014; Li et al., 2014; Lu et al., 2015; Zhao et al., 2018)。

通过泛基因组分析全面发掘群体基因组中的序列和结构变异，能够鉴定其中与有利表型相关的变异位点，为发掘和研究新的功能基因提供了重要依据。例如，利用 66 个水稻高质量基因组构建了水稻的泛基因组，从中共鉴定到 16 563 789 个 SNPs、5 549 290 个 Indels 和 933 489 个 SVs，分析了其中与开花时间相关的基因 *Hd3a* (Heading date 3a)、抗寒性基因 *COLD1* (Chilling tolerance divergence 1)、谷物重量基因 *GW6a* (Grain weight 6a)、分蘖角度基因 *TAC1* (Tiller Angle Control 1)、植株高度基因 *Sd1* (Semi dwarf 1)在不同材料间的遗传变异，表明 SNPs 变异是导致这些基因变异的基础(Zhao et al., 2018)。而利用 29 个高质量基因组构建的大豆泛基因组图谱，共鉴定获得 14 604 953 个 SNPs、12 716 823 个 Indels 和 776 399 个 SV（包含 723 862 个 PAVs、27 531 个 CNVs、21 886 个移位和 3 120 个倒置），发现有些结构变异在重要农艺性状调控中发挥重要作用，如 PAV、基因融合和 Indels 分别对种皮亮度、种皮颜色的驯化、缺铁失绿等性状具有重要影响（Liu et al., 2020）。

chinaXiv:202104.00064v1

ChinaXiv合作期刊

chinaXiv:202104.00064v1

同时，在不同层次上发现的多个序列和结构变异，不仅提供了更加丰富的变异信息，也为研究基因功能变异提供了更多素材。例如，通过六倍体普通小麦物种基因组间和亚基因组间的共线性分析，提出其“4A-5A-7B 染色体重排”是两次染色体易位事件的结果，并明确了重排的基因组区间的精细边界；并且在微观尺度上探讨了小麦春化基因 *Vrn2* (Vernalization2)的复杂进化历史，发现 *Vrn2* 同源基因在普通小麦基因组中的复杂分布是包含串联重复、多倍化、染色体易位和基因丢失在内的一系列事件叠加的结果(Chen et al., 2020)。另有研究利用 100 个番茄基因组中捕获到 238 490 个 SVs，构建得到泛结构变异(panSV)图谱，研究表明 SVs 是许多转座子的基础，而且 SVs 集中区域的基因渐渗现象严重，且群体中 90%的 SVs 变异可在泛基因组图谱中获得验证(Alonge et al., 2020)。

3 泛基因组学研究中功能基因的变异与多态性

遗传结构变异通常会导致基因功能的改变，泛基因组研究能够通过全面整合相关基因的遗传信息，揭示基因重组、融合等事件导致基因功能的获得、丢失，以及发掘新基因。例如，大豆缺铁萎黄病有关的候选基因被定位于 14 号染色体上，通过泛基因组研究发现该候选基因有两种单倍型：品种“中黄 13”所属的单倍型主要分布在低纬度地区；品种“威廉 82”所属的单倍型主要分布在高纬度地区，能够在高 PH 值、铁为不易吸收的难溶氧化物等环境中生存，这种单倍型启动子区有 1.4 kb 的 Indel 和外显子区有 5 个变异位点(Liu et al., 2020)。在油菜中通过全 PAV- GWAS (genome wide association study)分析发现 3 个开花抑制因子 *BnaA10.FLC*、*BnaA02.FLC* 和 *BnaC02.FLC* 的 PAVs 与油菜的开花时间和生态型分化密切相关，其中：冬油菜品种的 *BnaA10.FLC* 启动子区都含有 MITE (Miniature inverted repeat transposable element)插入；85%春油菜品种的 *BnaA10.FLC* 第一个外显子中含有 LINE (Long interspersed nuclear elements)插入；81%半冬性油菜品种的 *BnaA10.FLC* 启动子区含有 hAT 插入。表明 *BnaA10.FLC* 决定了油菜生态类型，是控制油菜开花的关键基因(Song et al., 2020)。

生物的表型往往是来自多个基因网络调控的结果，其中很多基因可能又同时对多个不同的表型性状具有影响，因此对某个表型的有利基因亦有可能对另一个表型具有不利影响。例如，现代番茄中的产量相关性状调控机制复杂，对 100 个番茄基因组的泛结构变异(pan-SV)的研究发现，由四个结构变异导致形成了三个 MADS-Box 基因，共同影响番茄的经济性状。其中 $j2^{TE}$ 基因型具有便于收获的无关节花梗表型，而 $ej2^{w}$ 基因型具有防止撞伤的大花萼表型，但两个基因型同时出现($j2^{TE}$ $ej2^{w}$)则会出现花序分枝过多而导致低育性的现象；*sb1*(Suppressor of branching 1)基因型能有效克服双隐性基因型的负面作用，实现增产。另外，*sb1* 基因型的表达可能受 1 号染色体上 *STM3* 基因的串联重复序列影响，且串联重复的拷贝数具有剂量效应 (Alonge et al., 2020)。因此，通过在更广泛的群体中研究基因功能变异对表型的影响，将有助于更加准确地对功能基因-表型的关联做出全面详细地评估，从而更好地的指导分子育种工作来培育出抗病性更强、产量更高、保质期更长、风味更好的作物品种，同时又不牺牲其他所期望的表型性状。作物泛基因组学研究已经发现了大量农艺表型与特定基因的存在、缺失和变异之间的多样化的相关性(Tao et al., 2019)，通过在泛基因组完整遗传图谱的基础上进行研究，将有利于彻底澄清其内在关联和相应的机理。

4 泛基因组学研究在种群遗传多样性和系统进化研究中的应用

对泛基因组学的研究，不仅可以全面地从基因组水平分析物种内遗传多样性，探究个体间的系统发生关系和表型差异的遗传基础，而且可以从物种、亚种水平分析基因组的序列变异和系统进化特征，为研究物种的起源及演化等重要生物学问题提供依据。例如，通过水稻

ChinaXiv合作期刊

泛基因组对 6 个水稻群体中与驯化有关的 7 个基因位点开展进化分析，发现 Aus 群体（Indica 的一个亚类群）并未全部聚在栽培稻进化分支上，从而提出 Aus 水稻群体处于不完全驯化选择状态(Zhao et al., 2018)。利用小麦泛基因组对 19 个小麦个体基因的 PAVs 进行了发掘并构建了系统进化树，发现小麦品种中国春位于进化树的基部，为小麦不同类型种质的系统进化关系和研究利用提供了理论依据(Montenegro et al., 2017)。而对 32 只乌鸦群体的泛基因组研究，将鸦属(*Corvus*)分为 Jackdaw 和 Crow 两大支系，并在此基础上探讨了不同进化分支上乌鸦的基因组结构变异和功能性状，尤其是发现乌鸦羽毛图案差异大，但遗传差异不大，主要受 *NDP* 基因上游 20 kb 处有一个大小为 2.25 kb 的 LTR (long terminal repeats)逆转座子插入调控 (Weissensteiner et al., 2020)。

泛基因组研究还可运用于对不同生态地理类型中差异较大的种质资源进行基因组测序，挖掘物种中新的基因，为候选基因的补充、物种多样性及适应性进化、起源经历和外来物种入侵性等问题的研究提供重要信息。例如，大豆群体的生物地理分析发现现代栽培大豆起源于中国的华北地区(Liu et al., 2020)，而水稻群体的相关研究发现现代栽培稻起源地应该包括中国华南地区(Huang et al., 2012)。此外，由于一些作物的基因库中包括多个物种，特别是具有不同遗传结构的野生近缘物种，需要构建含该作物所有品种及其近缘种的遗传图谱以进行更广泛的研究，因此也有学者提出了超-泛基因组(Super-Pangenome)的概念，以探讨更大范围种质群体的遗传基础及其多样性(Khan et al., 2019)。

5 泛基因组学研究的发展前景

真核生物的全部基因组信息包括核基因组、线粒体基因组和质体基因组。目前的泛基因组学研究大多关注的是核基因组，而线粒体和质体这两种细胞器的泛基因组研究也逐渐开始被重视。例如，研究者利用 PCAWG (The Pan-Cancer Analysis of Whole Genomes) 数据库中 2,658 个癌症样本及其匹配的正常组织样本的全基因组数据构建了人类线粒体基因组最全面的突变蓝图，并确定了多个高度突变类型，其中截断突变(truncated mutations)在肾脏癌症、结直肠癌和甲状腺癌中明显富集，提示了激活特殊的信号通路或会带来致癌影响(Yuan et al., 2020)。此外，有研究者利用 321 个辣椒的叶绿体基因组，构建了辣椒 5 个栽培种及 2 个变种的叶绿体泛基因组，其不但用系统发育信号分析揭示了辣椒属不同种间亲缘关系的远近，也对 7 个叶绿体泛基因组的 CDS (Coding sequence)、内含子和基因间隔区的遗传多样性进行了详尽分析，确定了 *rpl23* 和 *trnI* 的基因间隔区包含 44 bp 串联重复以及其他插入缺失和单核苷酸等丰富的变异(Magdy et al., 2019)。

在某些物种中，由于其基因组较大和可移动元件的比例较高等原因，使得泛基因组研究难以有效开展，因此，关注全部 RNA 信息的泛转录组(pan-transcriptome)研究开始逐渐兴起。许多重要作物如玉米(Hansey et al., 2012; Hirsch et al., 2014; 简银巧等, 2017)和大麦(Ma et al., 2019)，以及模式生物拟南芥(Gan et al., 2011)等的泛转录组研究均已有报道。

随着多种测序技术的结合和分析策略的发展，泛基因组学相关研究呈现爆发式增长，但是大多数研究的深入程度不一，许多数据结果仍有进一步深入挖掘的空间。尤其是构建完整的基因图谱后，很多研究止步于对某几个基因的结构变异进行鉴定，未进一步开展系统的功能研究，更不用说应用于生产实践。此外，随着大量生物信息学数据的积累，单个团队面对浩大的数据库也只能选择部分数据结果进行深入研究，难以充分利用现有的数据。例如，人类基因组计划从开始启动到现在已经过去 30 年，仍需大量的人力投入和研究分析去解决更多的问题。因而，完善的数据共享机制和良好平台是泛基因组学研究良性发展和应用的一个

chinaXiv:202104.00064v1

ChinaXiv合作期刊

重要条件。目前，我国已建立了国家基因组科学数据中心（NGDC, National Genomics Data Center），某些重要农作物或农业动物物种的泛基因组数据也建立了数据分享平台，如猪的泛基因组数据库 PIGPAN (http://animal.nwsuaf.edu.cn/code/index.php/pan-Pig)、大白菜基因组数据库 BRAD (the *Brassica* database, http://brassicadb.cn)和油菜泛基因组资源数据库(http://cbi.hzau.edu.cn/bnapus/)等。

此外，进一步整合更广泛的多层次群体基因组数据，如不同世代之间的泛基因组研究、整合多个物种的超-泛基因组研究等，可能是值得进一步探索的新方向(图 2)。另一方面，随着测序技术的不断发展，尤其是单细胞测序技术的发展，和测序成本进一步降低，单细胞分辨率的转录组图谱已经逐步开始在水稻和玉米的根发育研究中获得应用（Satterlee et al., 2020; Liu et al., 2021）。因此，同一个体不同组织器官的泛基因组或泛转录基因组研究，乃至不同细胞之间的泛基因组或泛转录基因组研究也可能成为新的发展方向(图 2)。

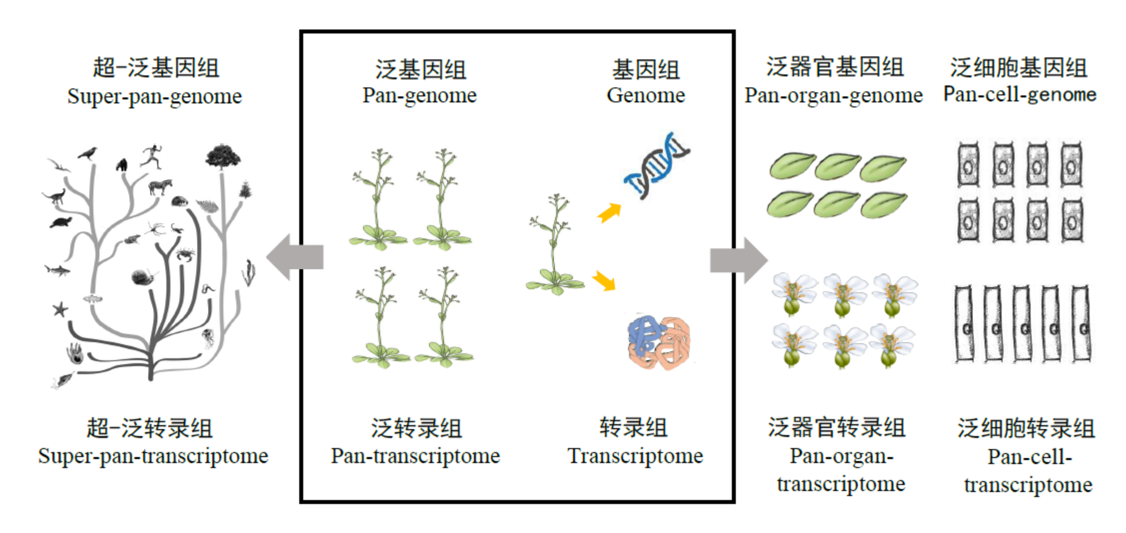


图 2 泛基因组（转录组）研究现状与可能的发展方向

Fig. 2 The research status and prospect of Pan-genome (Pan-transcriptome)

参考文献：

1000 GENOMES PROJECT CONSORTIUM, 2012. An integrated map of genetic variation from 1092 human genomes [J]. Nature, 491: 56–65.

ALONGE M, WANG X, BENOIT M, et al., 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato [J]. Cell, 182: 145–1161.

BAYER PE, GOLICZ AA, TIRNAZ S, et al., 2019. Variation in abundance of predicted resistance genes in the *Brassica oleracea* pangenome [J]. Plant Biotechnol J, 17: 789–800.

CHEN YM, SONG WJ, XIE XM, et al., 2020. A collinearity-incorporating homology inference strategy for connecting emerging assemblies in Triticeae tribe as a pilot practice in the plant pangenomic era [J]. Mol Plant, 13: 1694–1708.

DOMÍNGUEZ M, DUGAS E, BENCHOUAIA M, et al., 2020. The impact of transposable elements on tomato diversity [J]. Nat Comm, 11: 4058.

FU J, QIN Q, 2012. Pan-genomics analysis of 30 *Escherichia coli* genomes [J]. Hereditas, 34: 765–772. [付静，秦启伟，2012．30 株大肠杆菌的泛基因组学特征分析[J]．遗传，34：765–772．]

GAN X, STEGLE O, BEHR J, et al., 2011. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana* [J]. Nature, 477: 419–423.

GAO L, GONDA I, SUN H, et al., 2019. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor [J]. Nat Genet, 51: 1044–1051.

GOLICZ AA, BAYER PE, BARKER GC, et al., 2016. The pangenome of an agronomically

chinaXiv:202104.00064v1

ChinaXiv合作期刊

important crop plant *Brassica oleracea* [J]. Nat Comm, 7: 13390.
GORDON SP, CONTRERAS-MOREIRA B, WOODS DP, et al., 2017. Extensive gene content variation in the *Brachypodium distachyon* pangenome correlates with population structure [J]. Nat Comm, 8: 2184.
HANSEY CN, VAILLANCOURT B, SEKHON RS, et al., 2012. Maize (*Zea mays* L.) genome diversity as revealed by RNA-sequencing [J]. PLoS ONE, 7: e33071.
HIRSCH CN, FOERSTER JM, JOHNSON JM, et al., 2014. Insights into the maize pan-genome and pan-transcriptome [J]. Plant Cell, 26: 121–135.
HUANG XH, KURATA N, WEI XH, et al., 2012. A map of rice genome variation reveals the origin of cultivated rice [J]. Nature, 490: 497–501.
HÜBNER S, BERCOVICH N, TODESCO M, et al., 2019. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance [J]. Nat Plants, 5: 54–62.
HUFFORD MB, XU X, VAN HEERWAARDEN J, et al., 2012. Comparative population genomics of maize domestication and improvement [J]. Nat Genet, 44: 808–811.
HURGOBIN B, GOLICZ AA, BAYER PE, et al., 2018. Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus* [J]. Plant Biotechnol J, 16: 1265–1274.
JAYAKODI M, PADMARASU S, HABERER G, et al., 2020. The barley pan-genome reveals the hidden legacy of mutation breeding [J]. Nature, 588: 284–289.
JIAN YQ, 2017. Variations in pan-transcriptome and genome size in tropocal Maize (*Zea mays* L.) and Their Applications [D]. Beijing: Chinese Academy of Agricultural Sciences. [简银巧，2017．热带玉米全长泛转录组和基因组大小变异及应用[D]．北京：中国农业科学院.]
KHAN AW, GARG V, ROORKIWAL M, et al., 2019. Super-pangenome by integrating the wild side of a species for accelerated crop improvement [J]. Trends Plant Sci, 51: 1044–1051.
LI RQ, LI YR, ZHENG HC, et al., 2020. Building the sequence map of the human pan-genome [J]. Nature Biotechnol, 28: 57–63.
LI YH, ZHOU GY, MA JX, et al., 2014. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits [J]. Nature Biotechnol, 32: 1045–1052.
LIU Q, LIANG Z, FENG D, et al., 2021. Transcriptional landscape of rice roots at the single cell rsolution [J]. Mol Plant, 14:384–394.
LIU YC, DU HL, LI PC, et al., 2020. Pan-genome of wild and cultivated soybeans [J]. Cell, 182: 162–176.
LU F, ROMAY MC, GLAUBITZ JC, et al., 2015. High-resolution genetic mapping of maize pan-genome sequence anchors [J]. Nat Comm, 6: 6914.
MA YL, LIU M, STILLER J, et al., 2019. A pan-transcriptome analysis shows that disease resistance genes have undergone more selection pressure during barley domestication [J]. BMC Genomics, 20: 12.
MABIRE C, DUARTE J, DARRACQ A, et al., 2019. High throughput genotyping of structural variations in a complex plant genome using an original Affymetrix® axiom® array [J]. BMC Genomics, 20: 848.
MAGDY M, OU LJ, YU HY, et al., 2019. Pan-plastome approach empowers the assessment of genetic variation in cultivated *Capsicum* species [J]. Hort Res, 6: 108.
MEDINI D, DONATI C, TETTELIN H, et al., 2005. The microbial pan-genome [J]. Curr Opin Genet Dev, 15: 589–594.
MONTENEGRO JD, GOLICZ A, BAYER PE, et al., 2017. The pangenome of hexaploid bread wheat [J]. Plant J, 90: 1007–1013.
OU LJ, LI D, LV JH, et al., 2018. Pan-genome of cultivated pepper (*Capsicum*) and its use in gene presence-absence variation analyses [J]. New Phytol, 220: 360–363.
PINOSIO S, GIACOMELLO S, FAIVRE-RAMPANT P, et al., 2016. Characterization of the poplar pan-genome by genome-wide identification of structural variation [J]. Mol Biol Evol, 33: 2706–2719.
SATTERLEE JW, STRABLE J, SCANLON MJ, 2020. Plant stem cell organization and differentiation at single-cell resolution [J]. Proc Natl Acad Sci USA, 117: 33689–33699.
SCHATZ MC, MARON LG, STEIN JC, et al., 2014. Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of *aus* and *indica* [J]. Genome Biol, 15: 506.

chinaXiv:202104.00064v1

ChinaXiv合作期刊

SONG JM, GUAN ZL, HU JL, et al., 2020. Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus* [J]. Nat Plants, 6: 34–45.
SPRINGER NM, YING K, FU Y, et al., 2009. Maize inbreds exhibit high levels of copy number variation (CNV) and presence/absence variation (PAV) in genome content [J]. PLoS Genet, 5: e1000734.
SUN C, HU ZQ, ZHENG TQ, et al., 2017. RPAN: rice pan-genome browser for approximately 3000 rice genomes [J]. Nucl Acids Res, 45: 597–605.
TAO YF, ZHAO XR, MACE E, et al., 2019. Exploring and Exploiting Pan-genomics for Crop Improvement [J]. Mol Plant, 12: 156–169.
TETTELIN H, MASIGNANI V, CIESLEWICZ MJ, et al., 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: Implications for the microbial "pan-genome" [J]. Proc Natl Acad Sci USA, 102: 13950–13955.
TIAN XM, LI R, FU WW, et al., 2020. Building a sequence map of the pig pan-genome from multiple de novo assemblies and Hi-C data [J]. Sci Chin Life Sci, 63: 750–763.
VAN DE WEYER AL, MONTEIRO F, et al., 2019. A Species-Wide Inventory of NLR Genes and Alleles in *Arabidopsis thaliana* [J]. Cell, 178: 1260–1272.
WALKOWIAK S, GAO L, MONAT C, et al., 2020. Multiple wheat genomes reveal global variation in modern breeding [J]. Nature, 588: 277–283.
WANG WS, MAULEON R, HU ZQ, et al., 2018. Genomic variation in 3,010 diverse accessions of Asian cultivated rice [J]. Nature, 557: 43–49.
WANG YL, ZHU SS, YANG FS, et al., 2019. Pan-genome sequencing and comparative genomic analysis of atrazine-degrading bacteria [J]. Biotechnol Bull, 35: 90–99. [王娅丽，朱姗姗，杨峰山，等，2019. 莠去津降解菌泛基因组测序及比较基因组分析 [J]. 生物技术通报，35：90–99.]
WEISSENSTEINER MH, BUNIKIS I, CATALÁN A, et al., 2020. Discovery and population genomics of structural variation in a songbird genus [J]. Nat Comm, 11: 3403.
YAO W, LI GW, ZHAO H, et al., 2015. Exploring the rice dispensable genome using a metagenome-like assembly strategy [J]. Genom Biol, 16: 187.
YU JY, GOLICZ AA, LU K, et al., 2019. Insight into the evolution and functional characteristics of the pan-genome assembly from sesame landraces and modern cultivars [J]. Plant Biotechnol J, 17: 881–892.
YUAN Y, JU YS, KIM Y, et al., 2020. Comprehensive molecular characterization of mitochondrial genomes in human cancers [J]. Nat Genet, 52: 342–352.
ZHAO Q, FENG Q, LU HY, et al., 2018. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice [J]. Nat Genet, 50: 278.
ZHOU Y, CHEBOTAROV D, KUDRNA D, et al., 2020. A platinum standard pan-genome resource that represents the population structure of Asian rice [J]. Sci Data, 7: 113.
ZHU GT, HUANG SW, 2020. A 360-degree scanning of population genetic variations—A pan-genome study of soybean [J]. Chin Bull Bot, 55: 56–65. [祝光涛，黄三文，2020. 360度群体遗传变异扫描——大豆泛基因组研究[J].植物学报，55：403–406.]

chinaXiv:202104.00064v1

ChinaXiv合作期刊

chinaXiv:202104.00064v1

表 1 作物泛基因组相关研究

Table 1 Related researches on crop pan-genomes

研究对象 Object of study	样本数量 Sampling	主要研究内容 Main research content	参考文献 Reference
水稻（*Oryza sativa*，禾本科） Rice (*Oryza sativa*, Poaceae)	3 个水稻材料，包括 Nipponbare, IR64, DJ123 3 divergent rice including Nipponbare, IR64, DJ123	泛基因图谱构建 Pan-genome construction	Schatz et al., 2014
水稻（*Oryza sativa*，禾本科） Rice (*Oryza sativa*, Poaceae)	3 010 个亚洲栽培稻 3 010 diverse Asian cultivated rice	泛基因图谱构建和基因结构变异 Pan-genome construction and structural variation	Wang et al., 2018
水稻（*Oryza* spp.，禾本科） Rice (*Oryza* spp., Poaceae)	66 个水稻材料，包括栽培稻 *O. sativa* 和野生稻 *O. rufipogon* 66 divergent rice, including cultivated rice (*O. sativa*) and wild rice (*O. rufipogon*)	泛基因图谱构建、结构变异、功能基因变异和系统进化 Pan-genome construction, structural variation, functional gene variation and systematic evolution	Zhao et al., 2018
水稻（*Oryza sativa*，禾本科） Rice (*Oryza sativa*, Poaceae)	12 个水稻品种 12 of cultivated rice	泛基因图谱构建和基因结构变异 Pan-genome construction and structural variation	Zhou et al., 2020
玉米（*Zea mays*，禾本科） Maize (*Zea mays*, Poaceae)	75 个个体，包括野生种、本地品种和改良品种 75 wild, landrace and improved maize lines	基因结构变异、功能基因变异和系统进化 Structural variation, functional gene variation and systematic evolution	Hufford et al., 2012
玉米（*Zea mays*，禾本科） Maize (*Zea mays*, Poaceae)	503 个玉米自交系 503 maize inbred lines	泛转录组图谱构建和功能基因变异 Pan-transcriptome construction and functional gene variation	Hirsch et al., 2014
玉米（*Zea mays*，禾本科） Maize (*Zea mays*, Poaceae)	31 份热带玉米自交系 31 tropical maize inbred lines	泛转录图谱构建和序列(SNP)变异 Pan-transcriptome construction and sequence (SNP) variation	简银巧, 2017 Jian Q, 2017
玉米（*Zea mays*，禾本科） Maize (*Zea mays*, Poaceae)	440 个近交系、24 个高度重组近交系和 16 个 F1 代杂种 440 inbred lines, 24 highly recombinant inbred lines and 16 F1 hybrids	基因结构变异 Structural variation	Mabire et al., 2019
小麦（*Triticum aestivum*，禾本科） Wheat (*Triticum aestivum*, Poaceae)	中国小麦品种(Chinese Spring) 和 18 个小麦栽培种	泛基因图谱构建、基因结构变异和系统进化 Pan-genome construction, structural variation and systematic evolution	Montenegro et al., 2017
小麦（*Triticum aestivum*，禾本科） Wheat (*Triticum aestivum*, Poaceae)	15 个六倍体小麦，其中 10 个个体组装到染色体级别、5 个组装到 scaffold 级别 15 *Triticum aestivum* including 10 chromosome pseudomolecule and 5 scaffold assemblies of hexaploid wheat	泛基因图谱构建、基因结构变异和功能变异 Pan-genome construction, structural variation and functional gene variation	Walkowiak et al., 2020
大麦（*Hordeum vulgare*，禾本科） Barley (*Triticum aestivum*, Poaceae)	20 个大麦材料，包括了本地品种，栽培种和野生品种 20 varieties of barley comprising landraces, cultivars and a wild barley	泛基因图谱构建和基因结构变异 Pan-genome construction and structural	Jayakodi et al., 2020

ChinaXiv合作期刊

chinaXiv:202104.00064v1

		variation	
大豆（*Glycine soja*，豆科） Soybeans (*Glycine soja*, Fabaceae)	7 份代表性野生大豆，分布在中国北方、黄淮、南方和东北地区、日本、韩国和俄罗斯 7 *G. soja* representing the geographical adaptation within the species, distributed in North, Huanghuai and South regions of China, and Japan, Korea and Russia	泛基因图谱构建、基因结构变异、功能基因变异和系统进化 Pan-genome construction, structural variation, functional gene variation and systematic evolution	Li et al., 2014
大豆（*Glycine soja*，豆科） Soybeans (*Glycine soja*, Fabaceae)	26 份代表性大豆，包括 3 个野生大豆、9 个农家种和 14 个现代栽培品种，再加上已发表的中黄 13、Williams 82 和 W05 26 representative of soybeans, including 3 wild soybeans, 9 landraces, and 14 cultivars, and ZH 13, Williams 82 and W05 in previous studies	泛基因图谱构建、基因结构变异、功能基因变异和系统进化 Pan-genome construction, structural variation, functional gene variation and systematic evolution	Liu et al., 2020
番茄（*Solanum* spp.，茄科） Tomato (*Solanum* spp., Solanaceae)	725 个栽培番茄和野生番茄，栽培番茄分别为 372 个 SLL (*S. lycopersicum* var. *lycopersicum*)及 267 个 SLC (*S. lycopersicum* var. *cerasiforme*)；近亲 78 个 SP (*S. pimpinellifolium*) 和 8 个 SCG (*S. cheesmaniaeand* var. *galapagense*) 725 phylogenetically and geographically representative tomato, including 372 SLL, 267 SLC, 78 SP and 8 SCG.	泛基因图谱构建、基因结构变异和功能基因变异 Pan-genome construction, structural variation and functional gene variation	Gao et al., 2019
番茄（*Solanum* spp.，茄科） Tomato (*Solanum* spp., Solanaceae)	100 个番茄品种，包括 *S. pimpinellifolium*、*S. cheesmaniae*、*S. galapagense*、*S. lycopersicum* var. *cerasiforme* 和 *S. lycopersicum* 100 tomato including *S. pimpinellifolium*, *S. cheesmaniae*, *S. galapagense*, *S. lycopersicum* var. *cerasiforme* and *S. lycopersicum*	泛结构变异图谱构建和功能基因变异 Pan-structural-variation construction and functional gene variation	Alonge et al., 2020
辣椒（*Capsicum* spp.，茄科） Pepper (*Capsicum* spp., Solanaceae)	383 份辣椒材料，包括 355 个 *C. annuum*、4 个 *C. baccatum*、11 个 *C. chinense* 和 13 个 *C. frutescens* 383 cultivars, including 355 *C. annuum*, 4 *C. baccatum*, 11 *C. chinense* and 13 *C. frutescens*	泛基因图谱构建、基因结构变异和功能基因变异 Pan-genome construction, structural variation, and functional gene variation	Ou et al., 2018
辣椒（*Capsicum* spp.，茄科） Pepper (*Capsicum* spp., Solanaceae)	65 个个体，包括的物种有 *C. chacoense*、 *C. baccatum* var. *baccatum*、*C. baccatum* var. *pendulum*、*C. annuum* var. *annuum*、*C. annuum* var. *glabriusculum*、 *C. chinense* 和 *C. frutescens* 65 samples including *C. chacoense*, *C. baccatum* var. *baccatum*, *C. baccatum* var. *pendulum*, *C. annuum* var. *annuum*, *C. annuum* var. *glabriusculum*, *C. chinense* and *C. frutescens*	泛质体基因组图谱构建和基因结构变异 Pan-plastome construction and structural variation	Magdy et al., 2019
向日葵（*Helianthus annuus*，菊科） Sunflower (*Helianthus annuus*, Asteraceae)	493 份向日葵种质资源，包括 287 个栽培种品系、17 个美国原地方品种和 189 个野生近缘种 493 sunflower varieties including 287 cultivated lines, 17 Native American landraces and 189 wild accessions representing 11 compatible wild species.	泛基因图谱构建和基因功能研究 Pan-genome construction and functional gene variation	Hübner et al., 2019
甘蓝（*Brassica* spp.，十字花科） Cabbage (*Brassica* spp., Brasslcaceae)	9 种甘蓝品种(*B. oleracea*)和一种野生型近缘芸薹属物种(*B. macrocarpa*) 9 cultivated lines (*B. oleracea*) and one wild type (*B. macrocarpa*)	泛基因图谱构建和系统进化 Pan-genome construction and systematic	Golicz et al., 2016

chinaXiv:202104.00064v1

		evolution	
甘蓝（*Brassica oleracea*，十字花科） Cabbage (*Brassica* spp., Brasslcaceae)	同 Golicz et al., 2016 They used data of Golicz et al. (2016)	泛基因图谱构建、基因结构变异和功能基因变异 Pan-genome construction, structural variation and functional gene variation	Bayer et al., 2019
欧洲油菜（*Brassica napus*，十字花科） *Brassica napus* (Brasslcaceae)	53 个油菜品种，包括 33 个非人工合成系和 20 人工合成系。 53 *Brassica napus* varieties including 33 nonsynthetic accessions and 20 synthetic accessions	泛基因图谱构建和基因结构变异 Pan-genome construction and structural variation	Hurgobin et al., 2018
欧洲油菜（*Brassica napus*，十字花科） *Brassica napus* (Brasslcaceae)	8 个品种，包括 4 个半冬性油菜品种（中双 11、Gangan、浙油 7 号和胜利油菜）、2 个冬油菜品种（Tapidor 和 Quinta）、和 2 个春性油菜品种（Westar 和 No2127） Eight oilseed rape lines, including four SWORs (ZS11, Gangan, Zheyou7 and Shengli), two WORs (Tapidor and Quinta) and two SORs (Westar and No2127)	泛基因图谱构建、结构变异和功能基因变异 Pan-genome construction, structural variation and functional gene variation	Song et al., 2020
拟南芥（*Arabidopsis thaliana*，十字花科） *Arabidopsis thaliana* (Brasslcaceae)	64 个拟南芥个体 64 *A. thaliana*	泛 NLR 基因图谱构建、基因结构变异和功能进化 Pan-NLR-gene construction, structural variation and functional gene variation	Van de Weyer et al., 2019
杨树（*Populus* spp.，杨柳科） Poplar (*Populus* spp., Salicaceae)	3 个异交杨树（*P. nigra*、*P. deltoides* 和 *P. trichocarpa*） 3 intercrossable poplar species (*P. nigra*, *P. deltoides*, and *P. trichocarpa*)	泛基因图谱构建、基因结构变异和功能基因变异 Pan-genome construction, structural variation and functional gene variation	Pinosio et al., 2016
芝麻（*Sesamum indicum*，胡麻科） Sesame (*Sesamum indicum*, Pedaliaceae)	5 个芝麻品种，包括两个地方品种（Baizhima 和 Mishuozhima）和三个现在代培品种（Zhongzhi13、Yuzhi11 和 Swetha） 5 sesame varieties including 2 landraces (*S. indicum* cv. Baizhima and Mishuozhima) and 3 modern cultivars (*S. indicum* var. Zhongzhi13, Yuzhi11 and Swetha)	泛基因图谱构建和系统进化 Pan-genome construction and systematic evolution	Yu et al., 2019